

三 维 动 画 和 交 互 设 计 课 程 读 书 报 告



题目 基于深度学习的3D人脸与漫画建模草图系统

作者姓名 沈吴越

作者学号 21851094

指导教师 贝毅君

学科专业 软件工程

所在学院 软件学院

提交日期 二○一八年十二月

A Deep Learning Based Sketching System for 3D Face and Caricature Modeling

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: YiJun Bei

By

Wuyue Shen

Zhejiang University, P.R. China

2018

摘要

脸部建模一直是计算机图像和视觉领域的热门话题，包括卡通人物建模、人脸艺术设计、人脸实时重构等等，尤其是交互式人脸建模。

本文采用CNN神经网络构建了一个快速的、交互的、基于深度学习的人脸建模框架。通过简单勾勒人脸图画，本文的模型可以迅速生成对应的三维人脸模型，并且可以同时拟合面部轮廓和细节表情。同时我们也提供了多种方式进行快速的模型修改。本文的模型还适用于手势识别，用户可以通过手势对三维人脸模型进行快速的模型修改，便于用户操作。实验证明我们的结果具有高精度和快速度。

**关键词**：人脸模型，面部数据库，深度学习，人脸刻画，手势，素描模型

Abstract

Face modeling has been paid much attention in the field of visual computing.There exist many scenaios, including cartoon characters, avatars for social media, 3D face caricatures as well as face-related art and design.

In this paper, a fast, interactive, deep learning-based face modeling framework is constructed using CNN neural network. By simply sketching the face picture, the model of this paper can quickly generate the corresponding 3D face model, and can fit the facial contour and the detail expression at the same time. At the same time, we also provide a variety of ways to make rapid model changes. The model in this paper is also applicable to gesture recognition. Users can quickly modify the 3D face model by gestures, which is convenient for users to operate. Experiments have shown that our results are highly accurate and fast.

**Keywords：**Face Modeling, Face Database, Deep Learning, Face Caricatures. Gestures, Sketch-Based Modeling

目录

[摘要 3](#_Toc533612436)

[Abstract 3](#_Toc533612437)

[一、 论文简介 4](#_Toc533612438)

[二、 该文研究背景与问题描述 4](#_Toc533612439)

[2.1研究背景 4](#_Toc533612440)

[2.2 论文贡献 4](#_Toc533612441)

[三、 人脸模型数据准备 5](#_Toc533612442)

[四、 模型架构实现 6](#_Toc533612443)

[4.1 像素-输入层 6](#_Toc533612444)

[4.2造型-输入层 7](#_Toc533612445)

[4.3双线性输出层 7](#_Toc533612446)

[4.4顶点损失层 7](#_Toc533612447)

[五、 网络训练过程 8](#_Toc533612448)

[六、 实验与结果分析 9](#_Toc533612449)

[七、 思考与扩展 10](#_Toc533612450)

[7.1 AlexNet简介 11](#_Toc533612451)

[7.2 ReLU非线性函数 11](#_Toc533612452)

[7.3 局部响应归一化 12](#_Toc533612453)

[7.4 重叠池化 12](#_Toc533612454)

[7.5 采用dropout 13](#_Toc533612455)

[7.6 AlexNet整体结构 13](#_Toc533612456)

[参考文献 15](#_Toc533612457)

# 论文简介

论文题目：A Deep Learning Based Sketching System for 3D Face and Caricature Modeling，基于深度学习的3D人脸与漫画建模草图系统

发表于：ACM Transactions on Graphics，2017年10月

# 该文研究背景与问题描述

## 2.1研究背景

脸部建模一直是计算机图像和视觉领域的热门话题，包括卡通人物建模、人脸艺术设计、人脸实时重构等等，尤其是交互式人脸建模。

本文采用CNN神经网络构建了一个快速的、交互的、基于深度学习的人脸建模框架。通过简单勾勒人脸图画，本文的模型可以迅速生成对应的三维人脸模型，并且可以同时拟合面部轮廓和细节表情。同时我们也提供了多种方式进行快速的模型修改。本文的模型还适用于手势识别，用户可以通过手势对三维人脸模型进行快速的模型修改，便于用户操作。

## 2.2 论文贡献

1. 本文提出了一种用于三维人脸建模和漫画造型的新颖框架，这个框架具有高效建模的功能，三维人脸模型能够依据二维人脸素描的特征关系自动生成。同时该系统支持手势识别，用户可以使用手势对初始面部模型进行完善与修改。
2. 本文采用了一种基于深度回归网络的新颖卷积神经网络，正是该网络能够从二维人脸素描自动推断出其三维模型。该网络融合了卷积神经网络和二维素描的形状特征，具有两个独立的全连接层分枝。正是该结构保证了三维人脸模型建模的效率与准确性。
3. 生成了一个扩展的三维及二维的人脸数据库，该人脸数据库具有更多的身份，表情及不同程度的夸张。该数据库可直接用于训练与测试，方便后续人脸模型研究者的工作。

# 人脸模型数据准备

1. 面部表情扩充：原人脸数据集（Cao et al.2014）针对每一个人脸提供了20个面部表情，但许多面部表情间仅有镜像差异，这使得2D素描的特征提取变得难以进行与解释。为了提高模型的效果，本文针对原数据集中的11个面部表情进行了变形转换操作，并邀请艺术家重新设计了14个新的面部表情，使得面部表情数据更为丰富且多样。其中人脸的表情包括了人类基本的表情，例如：开心、难过和恐惧等。
2. 造型夸张：提取原有的人脸造型数据的梯度尺度因子，通过改变梯度尺度因子的改变，重构人脸的3D模型。但由于眼睛和嘴部的梯度不稳定，本文仅针对脸颊，下巴，后脑勺，鼻子和耳朵等部位进行夸张。同时设置了4个不同的夸张程度，夸张程度为1时造型和原有造型相同，夸张程度随等级上升而增加。丰富了人脸造型数据集。

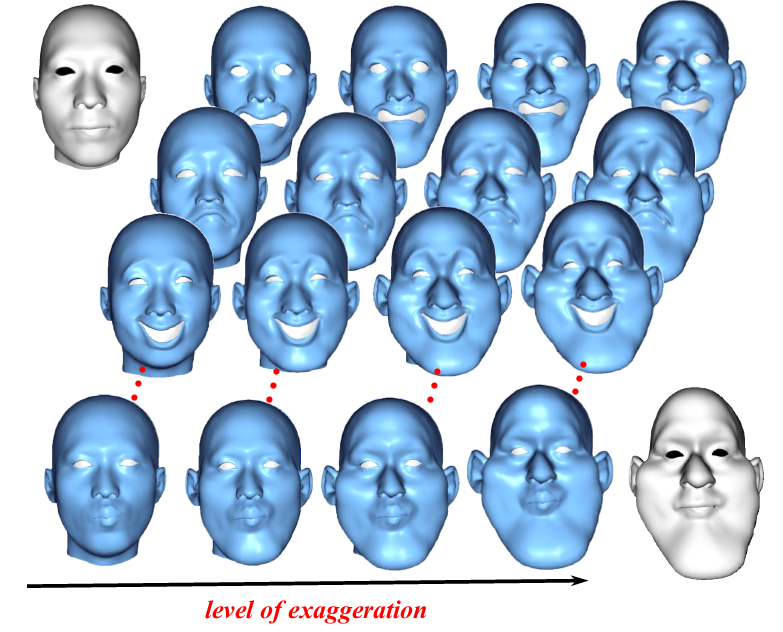


图1 面部表情夸张处理效果

1. 素描渲染：对原有的二维素描数据集进行了扩充，对原有的二维素描平面模型采用了数据增强技术，对眼睛、鼻子和嘴等部分进行了轮廓修改，使得二维素描平面模型的数据量得到了扩充，使得模型的泛化性能更佳。

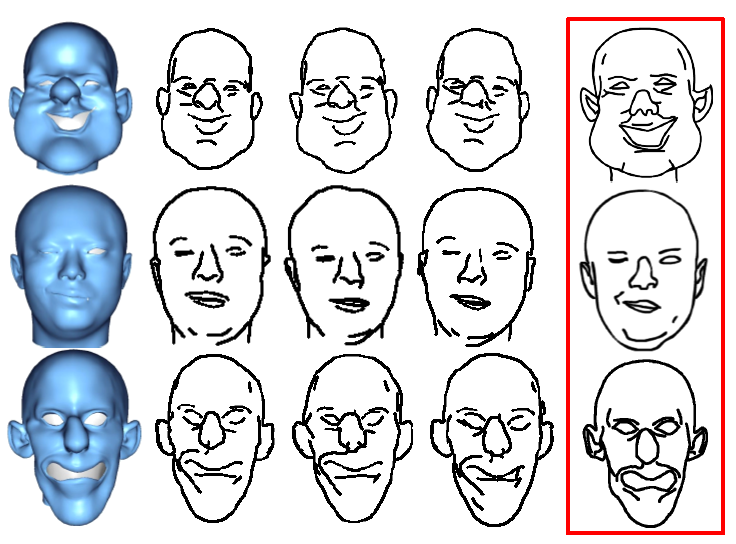


图2 面部表情轮廓修改处理效果

1. 手绘素描：为了使得二维素描平面模型更贴近真实手绘，本文采用了多位用户真实的手绘素描模型。

# 模型架构实现

## 4.1 像素-输入层

深度回归神经网络的输入是一个尺寸为的二值化图像，图像包含了轮廓，部位特征和线条特征等信息。该层的图像将会被投入AlexNet模块进行训练。该模块会提取二维素描图像中的特征，其共包括个卷积层，其尺寸分别为，最后由一个全连接层作为输出，并将特征信息输入到深度回归网络的全连接层中。

## 4.2造型-输入层

即使是最顶端的卷积层也只具备有限的视野而无法观察到整个图像，因此本文深层网络中的AlexNet模块不能够完全掌握输入草图中的全局上下文信息。为了充分利用草图中的信息，最大程度上得了解草图的特征，本层直接对轮廓和特征线上进行抽样，并抽取固定点数。通过另一个双线性表示这些采样点的2D位置模型，其中人脸身份和表情模式的系数仍然设置为50和16.这66个系数形成我们的造型-输入向量。在每一个分枝中，造型-输入层后都设置了一个具有512个神经元的全连接层，两个分支的输出最终作为输入，输入到双线性输出层中。

## 4.3双线性输出层

本文深度回归网络的输出是一组双线性人脸造型的系数，是最终能够重建三维人脸模型的3D顶点。如前文所述，这部分的双线性输出是包括两个部分：人脸身份和人脸表情，这两个部分将作为独立的模式，并赋予不同的权重和。为了避免这两个部分之间产生干涉，本文的深度神经网络具有两个不同的独立分枝，并具有不同数量的全连接层，来生成权重和。本文仍然借鉴了AlexNet的结构，生成权重μ的分枝包括个全连接层，而生成权重γ的分枝仅包含个全连接层。每个全连接层设置了个神经元，其数量是通过实际操作经验得到的，实验数据可表名该神经元数量效果最佳。

## 4.4顶点损失层

本文的深度回归网络是基于最小化顶点损失的训练模型。为了实现这一目标，本文将顶点损失定义为真实顶点与预测顶点之间的L2误差，所有的顶点都可以通过和矩阵得到。该L2误差定义为如下所示：

其中是对应于核心张量C的2D切片第个顶点，代表真实模型中第个顶点的位置，表示第个顶点的权重，是顶点的总数。可表示顶点作为某些顶点的相对重要性，例如那些在3D轮廓和特征线上的顶点具有更高的重要性。

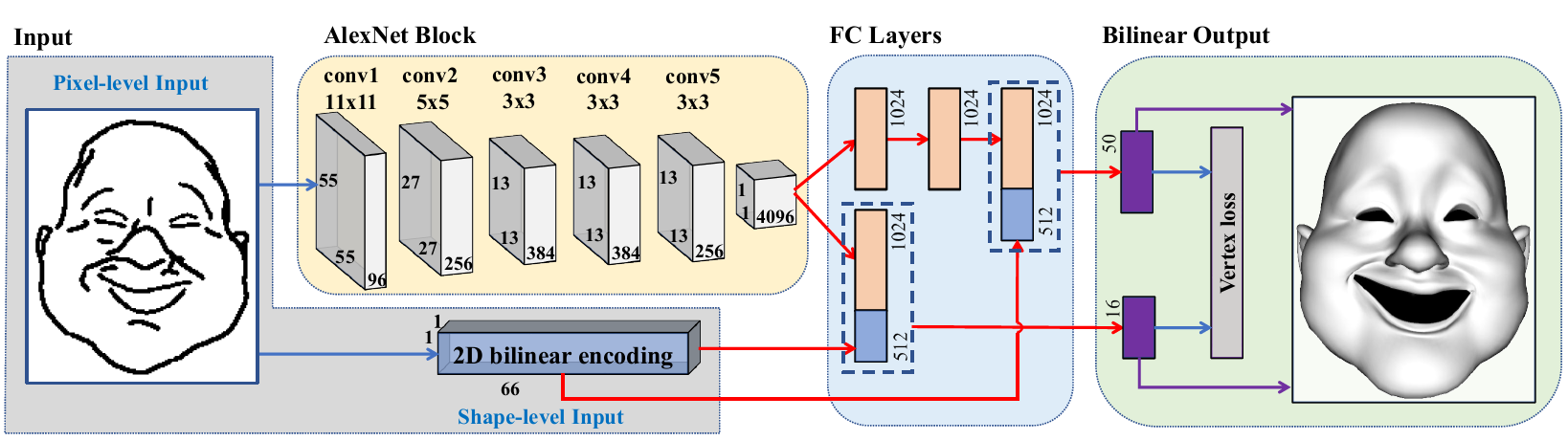


图3 本文所使用的网络框架

# 网络训练过程

整个神经网络的初始权重都是被设定为随机初始值，为了让模型的训练过程收敛得更加稳定，我们将训练过程分为三个步骤：分类器训练，和回归和最终调整。

在分类器训练这一环节中我们对分类器网络参数进行了随机初始化，两个分支的输出均采用了激活函数，由于扩展的数据集中具有个人脸身份和种面部表情，因此两个分支的输出层神经元数量分别设置为个和个。

第二个步骤中，我们通过两个分支预测出和两个顶点矩阵，这需要我们将softmax层替换为-回归层，该部分的神经元数量我们需要根据和参数的数量来设置。并计算出当前的L2损失值。

最后，我们将以L2损失为依据进行网络参数的调整，直至收敛。

为了使本文的网络更具泛化性能，本文对现有数据的顶点矩阵进行了随机插值，形成了新的顶点矩阵，并以此生成了新的人脸三维模型与平面素描模型加入到现有的数据库中。本文以新数据库的10%作为训练数据，其余数据作为训练集和验证集进行网络训练。并采用数据增强技术，在图像数据中引入噪声，对平面素描模型中的线条进行切除和形变。

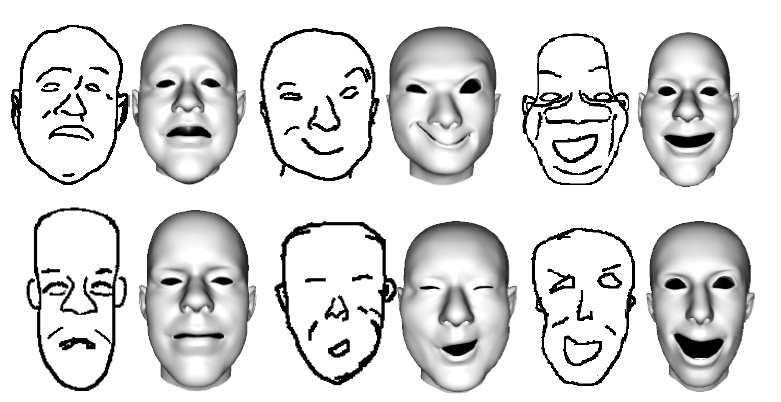


图4 训练后使用测试数据得到的部分结果

# 实验与结果分析

本文所有的网络搭建与训练均使用Caffe库（Jia et al.2014）实验环境如下：OS为Ubuntu16.04，GPU为GeForce Titan X，CPU为Inter Core i7 3.5GHz。

实验的参数设置：三轮训练过程的训练轮数分别为，学习率设置为，批量大小设置为,动量参数设置为，权重衰减设置为。下图为一系列具有不同形状与表情的3D面部造型，每个面部平均耗时8分钟完成建模。

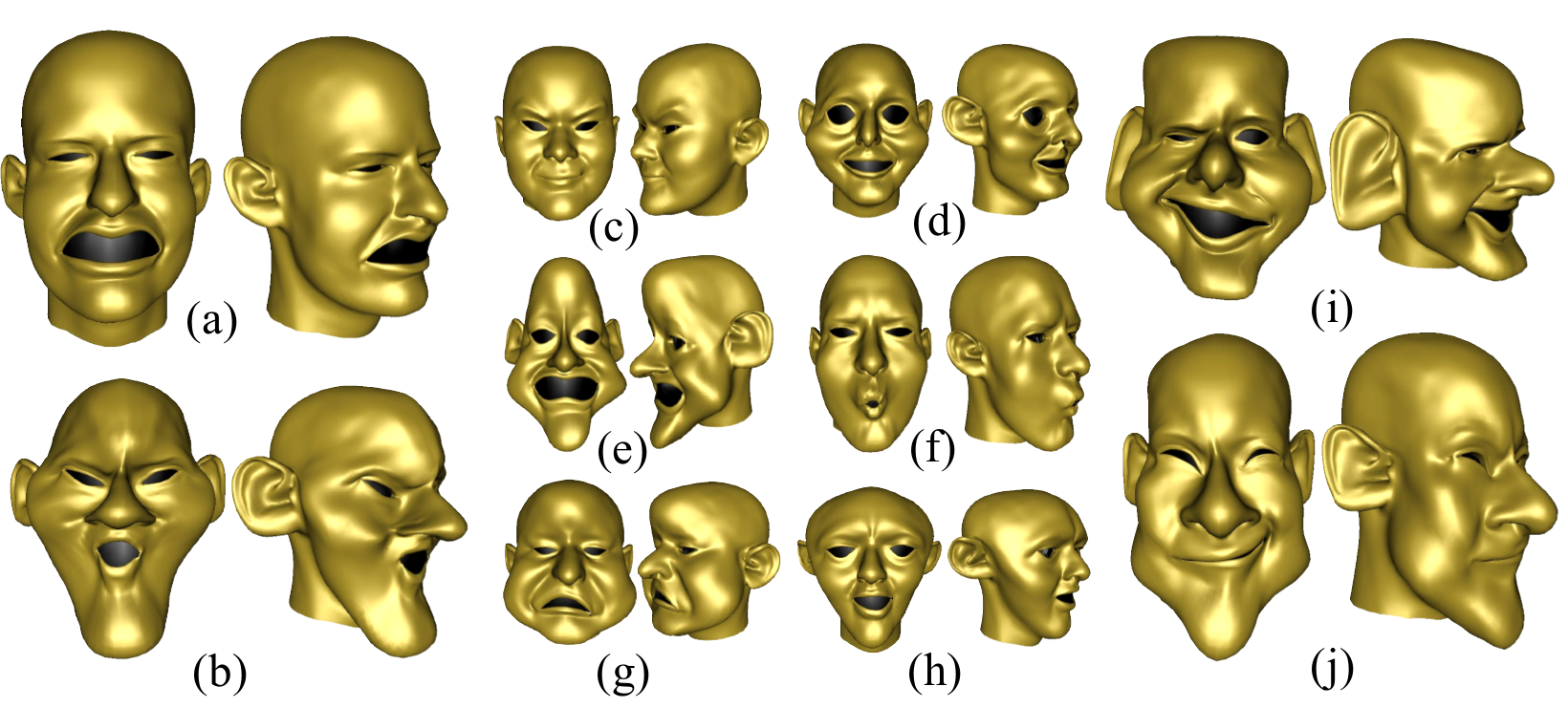


图5 本文的部分面部模型成果展示

本文的建模过程可以分为3个过程，初始2D素描模型绘制、再绘制与最终调整。初始绘制过程中用户会对自己想画的人脸造型进行简单的绘制，此时的模型较为接近手绘，因此存在线条不清晰，结构难以识别，面部特征不明显等一系列问题。本文的神经网络结构会根据该初始模型生成一个基本的三维人脸造型。如下图所示，此时生成的三维造型面部特征扔不是特别明显。当用户进入到再绘制阶段，本文的系统会仅保留原始素描模型中的关键线条，包括轮廓线，五官下巴等关键部位的特征线，更易于系统提出特征。此时用户可以修改现有的线条，使其更加符合用户心理预期的人脸造型，最终生成的模型见下图，可以看出人脸模型的特征得以细化，其造型也更接近用户的素描模型。下图显示了是否采用的模型差异。

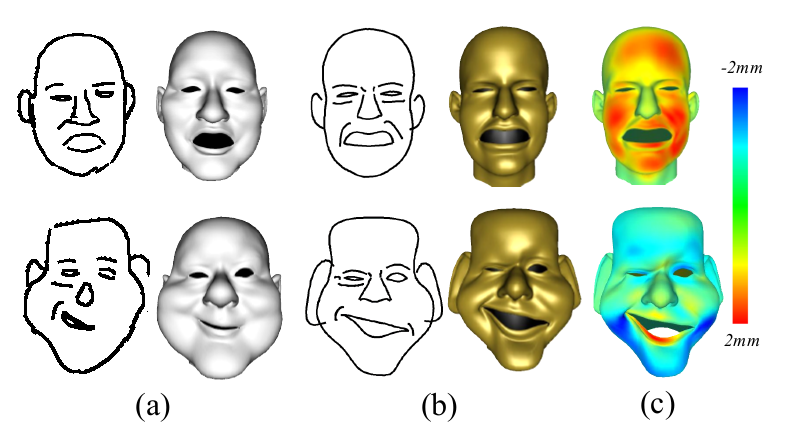


图5 初始绘制与再绘制结果

可以看出本文的系统可以将用户手绘的二维素描手绘模型转为三维面部模型，且面部表情及面部轮廓都达到了一个较为理想的效果。

# 思考与扩展

在阅读完本文后仍然会存在部分的疑问，例如对于AlexNet的结构与具体功能不是很了解，这和自己之前阅读的深度学习论文数量不够存在关系，因此也对本文中提到的许多专有名词和内容进行了更为深入的了解，本文深入了解了AlexNet的结构与创新点。

## 7.1 AlexNet简介

2012年AlextNet在ImageNet竞赛冠军后名声大噪，现如今大量的图像识别工作都借鉴了AlexNet的主要思想，AlexNet主要使用到的新技术点如下。

（1）成功使用ReLU作为CNN的激活函数，并验证其效果在较深的网络超过了Sigmoid，成功解决了Sigmoid在网络较深时的梯度弥散问题。虽然ReLU激活函数在很久之前就被提出了，但是直到AlexNet的出现才将其发扬光大。

（2）训练时使用Dropout随机忽略一部分神经元，以避免模型过拟合。Dropout虽有单独的论文论述，但是AlexNet将其实用化，通过实践证实了它的效果。在AlexNet中主要是最后几个全连接层使用了Dropout。

（3）在CNN中使用重叠的最大池化。此前CNN中普遍使用平均池化，AlexNet全部使用最大池化，避免平均池化的模糊化效果。并且AlexNet中提出让步长比池化核的尺寸小，这样池化层的输出之间会有重叠和覆盖，提升了特征的丰富性。

（4）提出了LRN层，对局部神经元的活动创建竞争机制，使得其中响应比较大的值变得相对更大，并抑制其他反馈较小的神经元，增强了模型的泛化能力。

## 7.2 ReLU非线性函数

AlexNet通常使用一个关于输入x的函数模拟神经元的输出f，这种标准函数是f(x)=tanh(x)或者f(x)=1/(（1+e^(-x)）)。在梯度下降训练时间上，这些饱和的非线性函数比不饱和非线性函数f(x)=max(0,x)更慢。使用ReLUs的深度卷积神经网络训练速度比同样情况下使用tanh单元的速度快好几倍。下图表示使用特定的四层卷积网络在数据集CIFAR-10上达到25%错误率所需的迭代次数。这个图表明如果使用传统的饱和神经元模型我们不可能利用这么大规模的神经网络对本文工作进行试验。

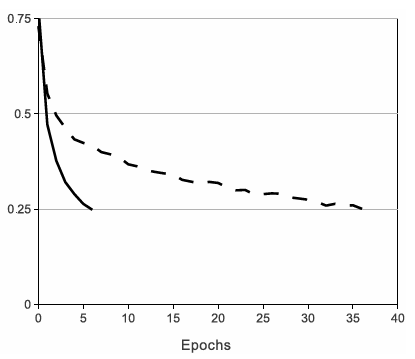
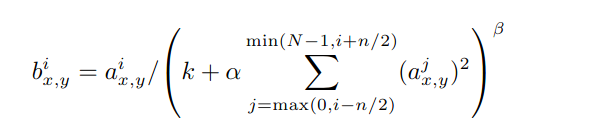


图6 传统激活函数与ReLu的收敛迭代次数对比

## 7.3 局部响应归一化

ReLUs具有符合本文要求的一个性质：它不需要对输入进行归一化来防止饱和。只要一些训练样本产生一个正输入给一个ReLU，那么在那个神经元中学习就会开始。但是，我们还是发现如下的局部标准化方案有助于增加泛化性能。a\_(x,y)^i

表示使用核i作用于(x,y)然后再采用ReLU非线性函数计算得到的活跃度，那么响应标准化活跃b\_(x,y)^i由以下公式计算出：



这里，对同一个空间位置的n个邻接核特征图（kernel maps）求和，N是该层的核的总数目。核特征图的顺序显然是任意的，并且在训练之前就已决定了的。这种响应归一化实现了侧抑制的一种形式，侧抑制受启发于一种在真实神经中发现的形式，对利用不同核计算得到的神经输出之间的大的活跃度生成竞争。常数k,n,α,β是超参数，它们的值使用一个验证集来确定。本文使用k=2,n=5,α=10-4,β=0.75。本文在某些特定的层中，采用ReLUs非线性函数后应用了该归一化。

## 7.4 重叠池化

CNN中的池化层归纳了同一个核特征图中的相邻神经元组的输出。通常，由邻接池化单元归纳的邻域并不重叠。更确切地说，一个池化层可以被看作是包含了每间隔S个像素的池化单元的栅格组成，每一个都归纳了以池化单元为中心大小为Z x Z的邻域。如果令S=Z，将会得到CNNs通常采用的局部池化。

## 7.5 采用dropout

结合多种不同模型的预测结果是一种可以降低测试误差的非常成功的方法，但是这对于已经要花很多天来训练的大规模神经网络来说显得太耗费时间了。但是，有一种非常有效的模型结合的方法，训练时间只需要原先的两倍。最新研究的技术，叫做“dropout”，它将每一个隐藏神经元的输出以50%的概率设为0。

这些以这种方式被“踢出”的神经元不会参加前向传递，也不会加入反向传播。因此每次有输入时，神经网络采样一个不同的结构，但是所有这些结构都共享权值。这个技术降低了神经元之间复杂的联合适应性，因为一个神经元不是依赖于另一个特定的神经元的存在的。因此迫使要学到在连接其他神经元的多个不同随机子集的时候更鲁棒性的特征。在测试时，本文使用所有的神经元，但对其输出都乘以了0.5，对采用多指数dropout网络生成的预测分布的几何平均数来说这是一个合理的近似。

AlexNet中前两个全连接层使用dropout。如果不采用dropout，其网络将会出现大量的过拟合。Dropout大致地使达到收敛的迭代次数增加了一倍。

## 7.6 AlexNet整体结构

如图所示，这个网络包含八个有权值的层：前五层是卷积层，剩下的三层是全连接层。最后一个全连接层的输出传递给一个1000路的softmax层，这个softmax产生一个对1000类标签的分布。本文的网络最大化多项Logistic回归结果，也就是最大化训练集预测正确的标签的对数概率。

第二、四、五层卷积层的核只和同一个GPU上的前层的核特征图相连。第三层卷积层和第二层所有的核特征图相连接。全连接层中的神经元和前一层中的所有神经元相连接。响应归一化层跟着第一和第二层卷积层。最大池化层，3.4节中有所描述，既跟着响应归一化层也跟着第五层卷积层。ReLU非线性变换应用于每一个卷积和全连接层的输出。

第一层卷积层使用96个大小为11x11x3的卷积核对224x224x3的输入图像以4个像素为步长（这是核特征图中相邻神经元感受域中心之间的距离）进行滤波。第二层卷积层将第一层卷积层的输出（经过响应归一化和池化）作为输入，并使用256个大小为5x5x48的核对它进行滤波。第三层、第四层和第五层的卷积层在没有任何池化或者归一化层介于其中的情况下相互连接。第三层卷积层有384个大小为3x3x256的核与第二层卷积层的输出（已归一化和池化）相连。第四层卷积层有384个大小为3x3x192的核，第五层卷积层有256个大小为 的核。每个全连接层有4096个神经元。

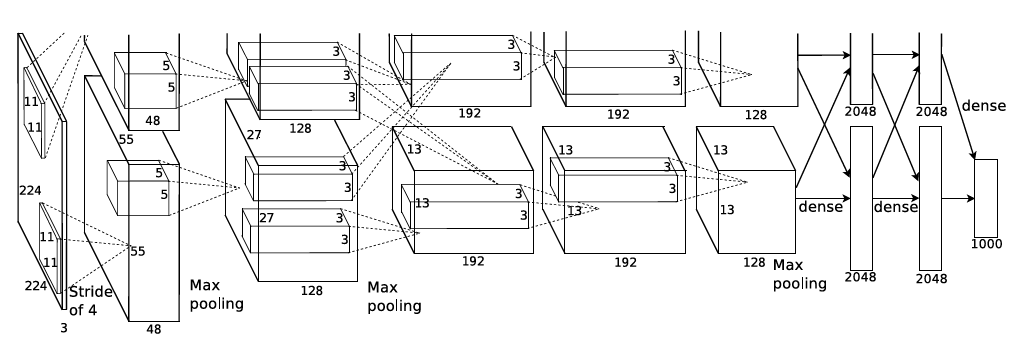


图7 AlexNet整体结构

# 参考文献

[1] Han X , Gao C , Yu Y . DeepSketch2Face: A Deep Learning Based Sketching System for 3D Face and Caricature Modeling[J]. Acm Transactions on Graphics, 2017, 36(4):126:1-12.

[2] Bae S H, Balakrishnan R, Singh K. ILoveSketch:as-natural-as-possible sketching system for creating 3d curve models[C]// Acm Symposium on User Interface Software & Technology. 2008.

[3] Clarke L, Chen M, Mora B. Automatic Generation of 3D Caricatures Based on Artistic Deformation Styles[J]. IEEE Trans Vis Comput Graph, 2011, 17(6):808-821.

[4] ImageNet Classification with Deep Convolutional Neural Networks Krizhevsky, Alex,Sutskever, Ilya,Hinton, Geoffrey E.-《Communications of the ACM》